下载APP

16 | 深度学习革命：深度学习推荐模型发展的整体脉络是怎样的？

2020-11-11 王喆

深度学习推荐系统实战 进入课程 >

讲述：王喆
时长 14:43 大小 13.49M

你好，我是王喆。今天，我们要开始学习激动人心的深度推荐模型部分了。

当下，几乎所有互联网巨头的推荐业务中，都有对深度学习推荐模型的落地和应用。比如说，从早期微软的 Deep Crossing、Google 的 Wide&Deep、阿里的 MLR，到现在影响力非常大的模型 DIN、DIEN，YouTube 的深度推荐模型等等。因此，对于我们算法工程师来说，紧跟业界的脚步，去了解和掌握深度学习推荐模型是非常必要的。

那你可能想问了，深度学习推荐模型这么多，发展这么快，也没有一个统一的模板，我们该学哪个，怎么学呢？我想说的是，算法工程师的工作是一个持续优化和迭代的过程，如果想要追求更好的推荐效果，我们的思路不应该只局限于某一个被成功应用的模型，而是应该把眼光放得更高、更宽，去思考这些成功的推荐模型在业界下一步的发展方向是什

么？有没有哪些其他的模型结构的思路可以借鉴。这些都是你在这个岗位上取得持续成功的关键。

那怎么才能做到这一点呢？我认为，只有建立起一个比较全面的深度学习模型知识库，我们才能在工作中做出正确的技术选择，以及为模型的下一步改进方向找到思路。

因此，这节课，我想和你深入聊一聊业界影响力非常大的深度学习推荐模型，以及它们之间的发展关系，带你从整体上建立起深度学习推荐模型的发展脉络。这不仅是我们建立行业知识储备的必需，也为后面实现深度推荐模型打下了基础。

深度学习对推荐系统的影响详解

在第一节课中，我们曾说过，**深度学习给推荐系统带来了革命性的影响**，能够显著提升推荐系统的效果，原因主要有两点，一是深度学习极大地增强了推荐模型的拟合能力，二是深度学习模型可以利用模型结构模拟用户兴趣的变迁、用户注意力机制等不同的用户行为过程。接下来，我们就结合这两点，来说说深度学习模型到强在哪里。

1．深度学习模型的强拟合能力

首先，我们来说说深度学习模型的强拟合能力。上一节课我们学习了经典的推荐算法，矩阵分解。在矩阵分解模型的结构（图 1 左）中，用户 One-hot 向量和物品 One-hot 向量分居两侧，它们会先通过隐向量层转换成用户和物品隐向量，再通过点积的方式交叉生成最终的打分预测。

但是，点积这种特征向量交叉的方式毕竟过于简单了，在数据模式比较复杂的情况下，往往存在欠拟合的情况。而深度学习就能大大加强模型的拟合能力，比如在 NeuralCF（神经网络协同过滤）模型中，点积层被替换为多层神经网络，理论上多层神经网络具备拟合任意函数的能力，所以我们通过增加神经网络层的方式就能解决模型欠拟合的问题了。

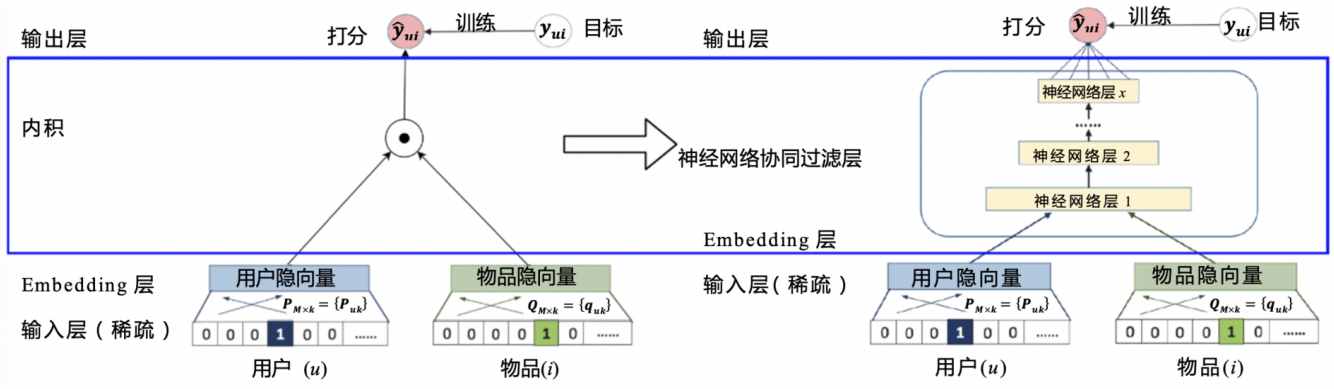

图1 矩阵分解模型示意图

如果你不知道什么是欠拟合、正确拟合和过拟合的现象，可能就无法理解神经网络到底解决了什么问题。这里我就带你看一张很经典的示意图，来详细聊聊这三种现象 。

“欠拟合”指的是模型复杂度低，无法很好地拟合训练集数据的现象。 就像图 2（左）展示的那样，模型曲线没办法“准确”地找到正负样本的分界线，而深度学习模型就可以大大增加模型的“非线性”拟合能力，像图 2（中）一样找到更加合适的分类面，更准确地完成分类任务。当然，过分复杂的深度学习模型存在着“过拟合”的风险，**“过拟合”是指模型在训练集上的误差很小，但在测试集上的误差较大的现象**，就像图 2（右）一样，模型曲线过分精确地刻画分界线而忽略了对噪声的容忍能力，这对于未知样本的预估来说往往是不利的。

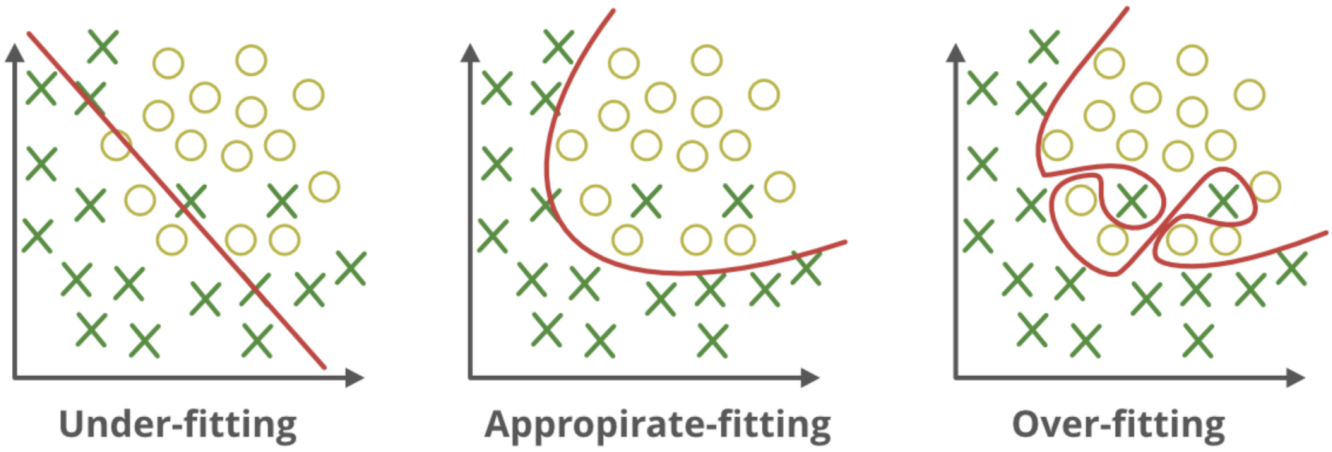

图2拟合能力示意图（图片来自Geeksforgeeks）

2. 深度学习模型结构的灵活性

说完了深度学习模型的强拟合能力，我们再来看看它的灵活性。这里，你可能会有疑问了，灵活性和深度学习模型模拟用户行为有什么关系呢？我们先接着往下看。

如果你读过一些深度学习相关的论文，肯定会发现，每篇论文中的模型结构都不尽相同，就像图 3 展示的那样，它们有的是好多层，有的像一个串是串在一起的，而有的像一张网，拥有多个输入和多个输出，甚至还有的像金字塔，会从输入到输出逐层变窄。

虽然，模型结构的复杂性让我们难以掌握它们的规律，但也正是因为深度模型拥有这样的灵活性，让它能够更轻松地模拟人们的思考过程和行为过程，让推荐模型像一个无所不知的超级大脑一样，把用户猜得更透。

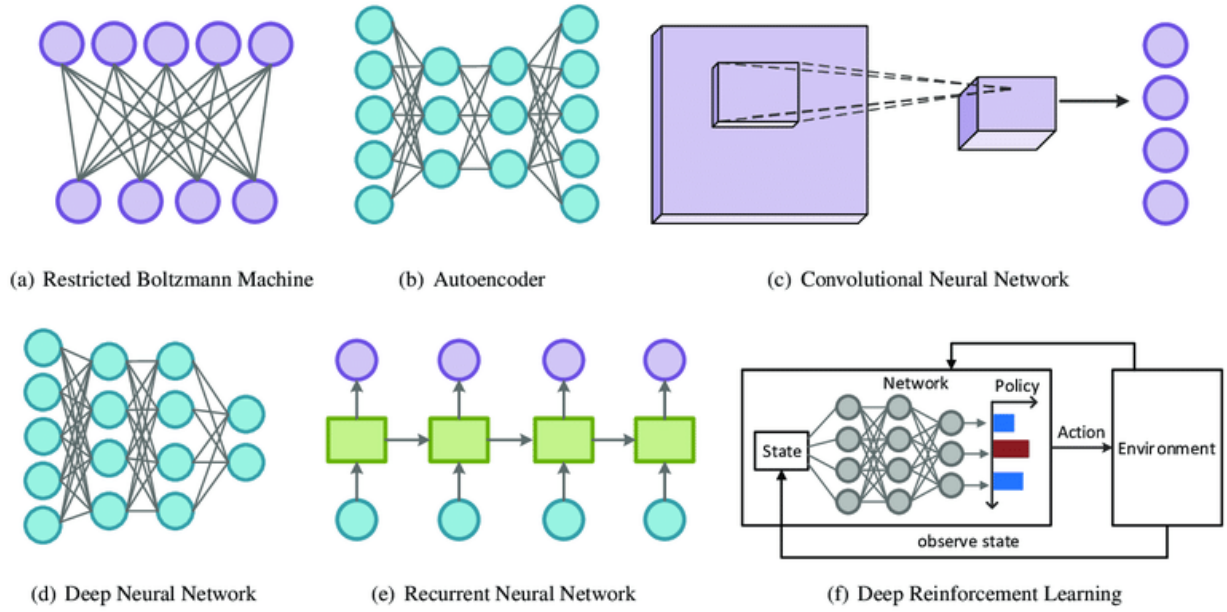

图3 不同深度学习模型的结构（图片来自ResearchGate）

这其中典型的例子就是阿里巴巴的模型 DIN（深度兴趣网络）和 DIEN（深度兴趣进化网络）。它们通过在模型结构中引入注意力机制和模拟兴趣进化的序列模型，来更好地模拟用户的行为。

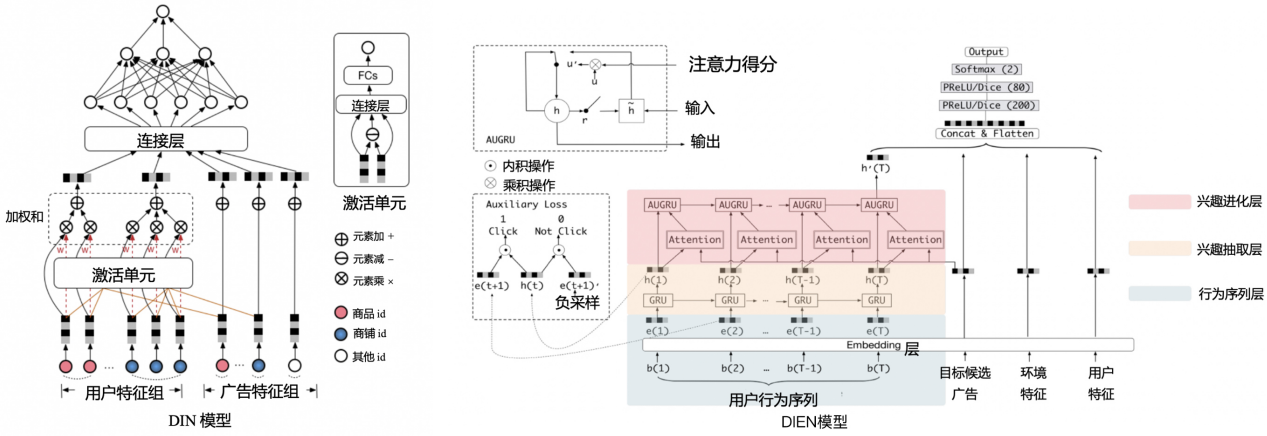

图4 DIN模型（左）和DIEN的模型（右）示意图

我们重点关注图 4 的 DIN 模型，它在神经网络中增加了一个叫做“激活单元“的结构，这个单元就是为了模拟人类的注意力机制。举个例子来说，我们在购买电子产品，比如说笔记本电脑的时候，更容易拿之前购买电脑的经验，或者其他电子产品的经验来指导当前的购买行为，很少会借鉴购买衣服和鞋子的经验。这就是一个典型的注意力机制，我们只会注意到相关度更高的历史购买行为，而 DIN 模型就是模拟了人类这样的注意力特点。

DIN 模型的改进版 DIEN 模型就更厉害了，它不仅引入了注意力机制，还模拟了用户兴趣随时间的演化过程。我们来看那些彩色的层，这一层层的序列结构模拟的正是用户兴趣变

迁的历史，通过模拟变迁的历史，DIEN 模型可以更好地预测下一步用户会喜欢什么。

这些通过改变模型结构来模拟用户行为的做法不胜枚举。很多重要的深度学习模型的改进动机正是基于这样的原理。也正是因为这样的灵活性，正确、全面地掌握不同深度学习模型的特点，以及它们之间的发展关系变得异常重要，只有这样，我们才能做到在实践中有的放矢、灵活应用。

深度学习推荐模型的演化关系图

说了这么多，我们到底该怎么掌握不同深度学习模型之间的关系呢？这里，我梳理出了一张深度学习模型 5 年内的发展过程图，图中的每一个节点都是一个重要的模型结构，节点之间的连线也揭示了不用模型间的联系。

接下来，我就带你梳理一下图中重要模型的原理，以及不同模型间的关系。而在之后的课程中，我们还会进一步学习重点模型的技术细节，并且基于 TensorFlow 对它们进行实现。

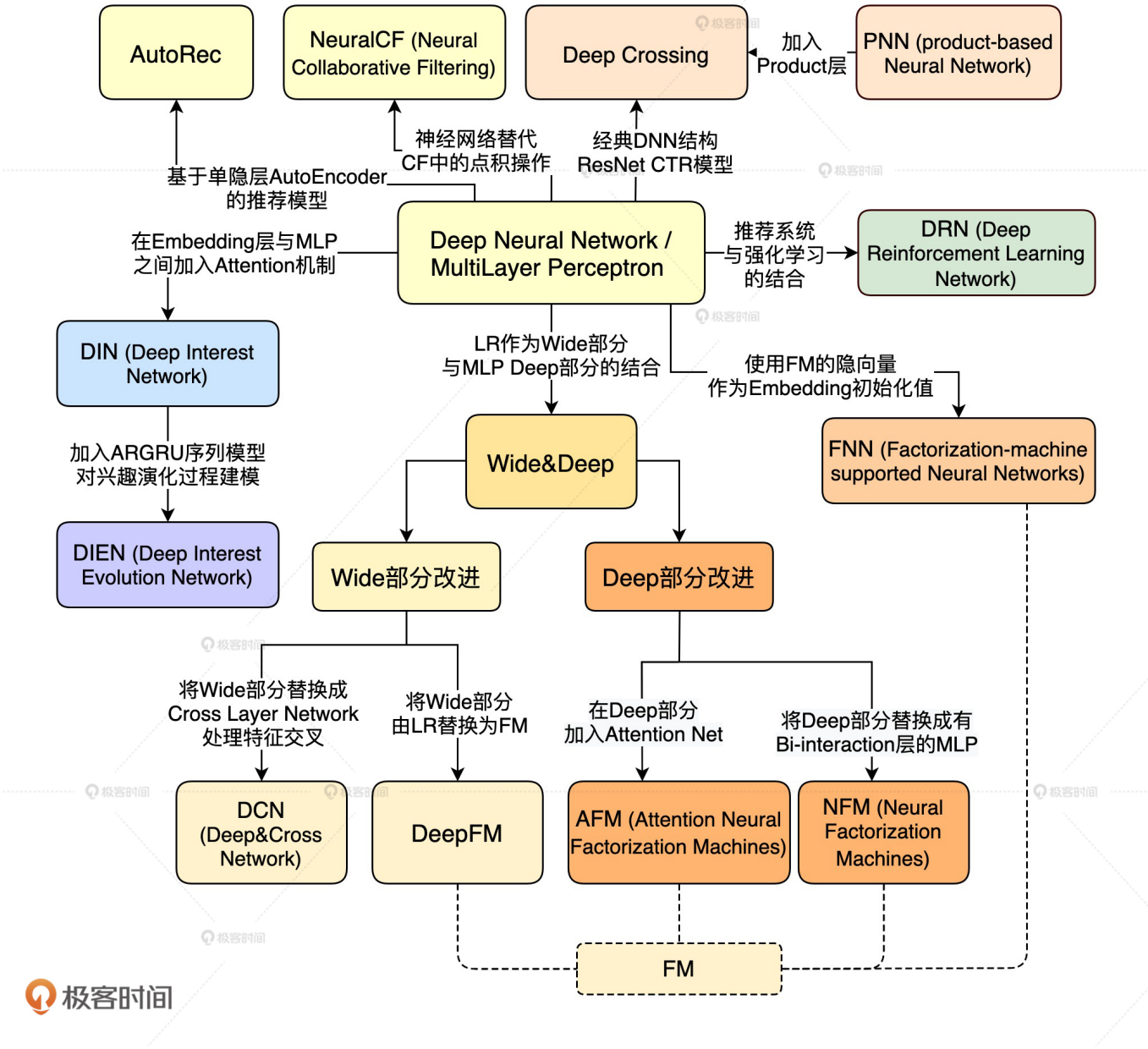

图5 主流深度学习推荐模型的演化图谱

首先，我们来看整个演化图最中心部分，这是深度学习最基础的结构，我们叫它“多层神经网络”或者“多层感知机”，简称 MLP（MultiLayer Perceptron）。多层感知机的原理我们在第 3 节课中讲过，它就像一个黑盒，对输入的特征进行深度地组合交叉，然后输出对兴趣值的预测。其他的深度推荐模型全都是在多层感知机的基础上，进行结构上的改进而生成的，所以**“多层感知机”是整个演化图的核心**。

从多层感知机向上，还有一个重点模型我们需要知道，那就是 Deep Crossing。Deep Crossing 实际上是一类经典深度学习模型的代表，相比于 MLP，Deep Crossing 在原始特征和 MLP 之间加入了 Embedding 层。这样一来，输入的稀疏特征先转换成稠密 Embedding 向量，再参与到 MLP 中进行训练，这就解决了 MLP 不善于处理稀疏特征的

问题。可以这么说，**Embedding+MLP 的结构是最经典，也是应用最广的深度学习推荐模型结构**。

从 MLP 向下，我们看到了 Google 提出的推荐模型 Wide&Deep。它把深层的 MLP 和单层的神经网络结合起来，希望同时让网络具备很好的“记忆性”和“泛化性”。对“记忆性”和“泛化性”这两个名词陌生的同学也不用着急，我们后面的课程会专门来讲解 Wide&Deep。

Wide&Deep 提出以来，凭借着“易实现”“易落地”“易改造”的特点，获得了业界的广泛应用。围绕着 Wide&Deep 还衍生出了诸多变种，比如，通过改造 Wide 部分提出的 Deep&Cross 和 DeepFM，以及通过改造 Deep 部分提出的 AFM，NFM 等等。总之，**Wide&Deep 是业界又一得到广泛应用的深度推荐模型**。

除此之外，我们还可以看到经典的深度学习模型跟其他机器学习子领域的交叉。这里，我给你举 3 个比较著名的例子：第 1 个是深度学习和注意力机制的结合，诞生了阿里的深度兴趣网络 DIN，浙大和新加坡国立提出的 AFM 等等；；第 2 个是把序列模型引入 MLP+Embedding 的经典结构，诞生了阿里的深度兴趣进化网络 DIEN；第 3 个是把深度学习和强化学习结合在一起，诞生了微软的深度强化学习网络 DRN，以及包括美团、阿里在内的非常有价值的业界应用。

看了这诸多模型的演进过程，你肯定想问模型的演化有什么规律可循吗？接下来，我就把我总结出的，关于模型改进的 4 个方向告诉你 。

一是改变神经网络的复杂程度。 从最简单的单层神经网络模型 AutoRec，到经典的深度神经网络结构 Deep Crossing，它们主要的进化方式在于增加了深度神经网络的层数和结构复杂度。

二是改变特征交叉方式。 这种演进方式的要点在于大大提高了深度学习网络中特征交叉的能力。比如说，改变了用户向量和物品向量互操作方式的 NeuralCF，定义了多种特征向量交叉操作的 PNN 等等。

三是把多种模型组合应用。 组合模型主要指的就是以 Wide&Deep 模型为代表的一系列把不同结构组合在一起的改进思路。它通过组合两种甚至多种不同特点、优势互补的深度学习网络，来提升模型的综合能力。

四是让深度推荐模型和其他领域进行交叉。 我们从 DIN、DIEN、DRN 等模型中可以看出，深度推荐模型无时无刻不在从其他研究领域汲取新的知识。事实上，这个过程从未停歇，我们从今年的推荐系统顶会 Recsys2020 中可以看到，NLP 领域的著名模型 Bert 又与推荐模型结合起来，并且产生了非常好的效果。一般来说，自然语言处理、图像处理、强化学习这些领域都是推荐系统经常汲取新知识的地方。

总的来说，深度学习推荐模型的发展快、思路广。但每种模型都不是无本之木，它们的发展脉络都有迹可循，而想要掌握好这些模型，在实际工作中做到拿来就用，我们就需要让这些模型脉络图像知识树一样扎根在心中，再通过不断地实践来掌握技术细节。

小结

这节课，我们通过学习深度学习对推荐系统的影响要素，以及经典深度学习模型之间的关系，初步建立起了深度学习模型的知识库。

我们知道，深度学习能够提升推荐系统的效果有两个关键因素，分别是它的“强拟合能力”和“结构的灵活性”。

对于“强拟合能力”来说，深度学习模型可以大大增加模型的“非线性”拟合能力，对复杂数据模型进行更准确的分类，避免“欠拟合”现象的发生，从而提升推荐效果。

对于“结构的灵活性”来说，深度学习模型可以通过灵活调整自身的结构，更轻松恰当地模拟人们的思考过程和行为过程，把用户猜得更透。

而整个深度学习推荐模型的演化过程，是从最经典的多层神经网络向不同方向开枝散叶，比如结合协同过滤发展出了 NerualCF，加入 Embedding 层发展出以 Deep Crossing 为代表的 Embedding+MLP 的结构，以及把深度神经网络和单层网络结合起来发展出 Wide&Deep 模型等等。

在这节课里，我们可以先忽略每个模型的细节，着重建立一个整体的知识框架，在之后的课程中，我不仅会带你一一揭晓它们的技术细节，还会利用 TensorFlow 实现其中几个经典的模型。期待继续与你一起学习！

最后，我还是把这节课的重点知识梳理成了表格的形式，你可以借助它来复习巩固。

知识点	知识描述
深度学习能提升推荐系统效果的原因	1.深度学习模型具备更强的数据拟合能力； 2.深度学习模型结构的灵活性使之更容易模拟用户思考和行为的过程
深度学习模型的出发点	多层神经网络 MLP
神经网络+协同过滤	NeuralCF
神经网络+Embedding	Deep Crossing
神经网络+乘积层	PNN
深度网络+浅层网络	Wide&Deep、DeepFM、DCN
神经网络+注意力机制	DIN、AFM
神经网络+序列模型	DIEN
深度学习+强化学习	DRN

课后思考

有的同学说，深度学习这么流行，我把一些经典的深度模型结构实现好，肯定能提升我们公司推荐系统的效果，你觉得这种观点有问题吗？你觉得除了模型结构，还有哪些影响推荐效果的因素？为什么？

好啦，关于深度学习模型的知识库你建立起来了吗？欢迎把你的疑问和思考分享到留言区，也欢迎你能把这节课转发出去，我们下节课见！

提建议

更多学习推荐

© 版权归极客邦科技所有，未经许可不得传播售卖。 页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

上一篇 15 | 协同过滤：最经典的推荐模型，我们应该掌握什么？

下一篇 模型实战准备（一） | TensorFlow入门和环境配置

精选留言 (5)

写留言

Dikiwi
2020-11-11

算法选型一定是与业务相关的，按照老师的说法就是没有银弹。比如阿里的购买序列是具有一定的所谓演化的概念的，但不适合feed信息流，feed的兴趣相对会比较稳定。但在结合业务特点情况下选择以后的模型优化好了是大概率有不错的效果提升的。

个人觉得深度学习可以根据业务特点灵活定制模型结构的是其最具有魔力的地方。

展开

作者回复: 说的太好了，关于信息流和电商的思考也很好，推荐其他同学参考。

4

Eio
2020-11-11

请问老师 深度模型中怎么没有xDeepFM模型呢 我看它在学术界的评价要比deepfm好 但是工业界综合各个方面的因素是不是效果并不理想呢

展开

作者回复: 没有特殊的原因，xDeepFM也非常好，我们没必要也没有空间把所有模型都覆盖到。因为xDeepFM也属于FM+DNN的一个衍生模型。

而且不要过度迷信这些paper里面的效果，外界的传言，自己在自己的数据上判断，用不同结构去做试验，得出最终solid的结论，做算法工程师有时候要自信一些，没有人比你自己更了解你要解决的问题。

3

Sicolas Flamel

2020-11-11

从数据输入看，对数据源的处理（特征选择）也是一个难点。尽管处理特征的能力在深度学习背景下变强，但依然需要将不同语义下的数据汇总到一起。
另外，从影响因素来说，用户购买行为依然有一部分随机因素难以捕捉到，或者说不同的人受影响的随机因素不同。有的人心情不好就喜欢买奇奇怪怪的东西，那机器有办法捕捉到心情的特征吗？有的人随着年龄见识的增长，频繁购买的东西类别也会发生变化。还...

展开

作者回复: 确实存在这个问题，但我要说的是既然是做机器学习，我们只能从统计的角度理解问题，这些special case确实会损伤模型效果，但没办法，有时候我们不能钻牛角尖。

当然，对于特征的处理我完全赞同，即使在深度学习背景下，特征的挑选，预处理仍然重要，数据和特征是上限，模型结构只是逼近这个上限。

2

那时刻

2020-11-12

老师提到深度学习模型可以大大增加模型的“非线性”拟合能力，对复杂数据模型进行更准确的分类，避免“欠拟合”现象的发生。请问对于过拟合，深度学习模型又怎么避免和处理呢？是通过处理超参数么？

作者回复: 过拟合说明深度学习模型过于复杂了，可以通过减少模型层数，每层的神经元数量，采用dropout，增加L1正则化项等方式解决。具体的方法篇幅原因不展开讲，推荐去搜一下就清楚了。

1

超~~

2020-11-11

老师你好，关于阿里的EGES，怎么通过训练得到的item embedding ，得出user的embedding？可以直接avg么，谢谢老师

展开

作者回复: 这个我们之前应该介绍过，item emb通过item2vec预训练，user embedding可以通过avg得到。

第二种方法是放到EGES中，加入item emb layer和user emb layer，做end2end训练。